WAIS promises easy text retrieval
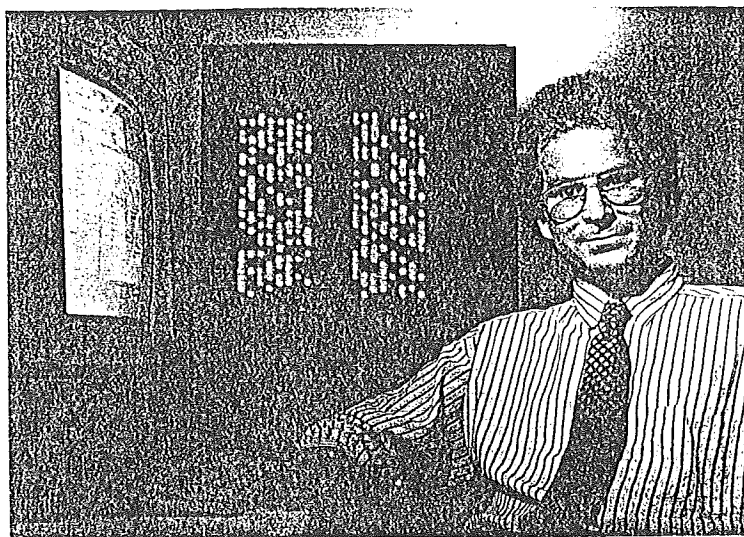
Prototype links Mac, Connection Machine

By Henry Norr

Cupertino, Calif. — Thinking Machines Corp., a pioneer in the development of high-powered parallel-processing supercomputers, has joined with Apple, Dow Jones & Co. and KPMG Peat Marwick to develop a new technology designed to simplify the retrieval of textual information stored in personal files, corporate records and remote databases.

Called the Wide Area Information Server (WAIS) project, the collaborative venture has been under way for almost two years. Peat Marwick recently completed a four-month experiment with the system, using WAIStation, a prototype Mac front end developed by Thinking Machines of Cambridge, Mass. Engineers from Apple's Advanced Technology Group have combined the WAIS technology with a custom interface to build a prototype personal electronic newspaper.

The WAIS project was designed in part to address problems caused by the proliferation of electronic data within large organizations.

"Corporations are starting to gag on gigabytes of word processing files, memos, reports, articles and E-mail archives," said Brewster Kahle, WAIS project leader for Thinking Machines. "Corporate memory is stored in this form, but executives have no easy way to get at it."

But the WAIS project was intended from the beginning, Kahle said, to be more than a traditional executive information system working only within corporate bounds. The objective was to lay the foundations for a scalable system that would allow users to tap a variety of data sources, including large commercial databases, through a uniform interface. Users, according to the plan, should be able to search for any available information without having to master the internal organization and query techniques of each source.

Brewster Kahle, WAIS project leader, helped develop an experimental text-retrieval system that can use a Thinking Machines supercomputer as a server and the Mac as a front end.

DRAKE SOREY

Peat Marwick tries 'partner-friendly' system

When officials of Thinking Machines Corp., Dow Jones & Co. and Apple first broached the concept of the Wide Area Information Server (WAIS) with KPMG Peat Marwick, representatives of the accounting giant were intrigued but cautious, according to Brewster Kahle, project leader for Thinking Machines.

They weren't interested, he said, in another complex querying application that busy tax consultants, accountants and managers would never bother to use. But they agreed to participate in the project, according to Kahle, on the promise of a system that would be genuinely "partner-friendly," with "no algebra — no ifs, ands or buts."

After a year of preliminary work, an experimental WAIS R&D project went on-line at Peat Marwick last October. About 10 users at the company's Montvale, N.J., headquarters, including "very senior partners," took part in the experiment, along with two others in Manhattan and 10 more on the West Coast, according to Robin Palmer, senior manager and WAIS project leader at KPMG Peat Marwick in San Jose, Calif. The remote users were connected by leased lines to a WAIS server running on a Connection Machine, a Thinking Machines parallel-processing system, installed in Montvale.

The Peat Marwick experiment relied on WAIStation, a Mac-based client software program developed by Thinking Machines, as a front end. To prepare a query, users need only enter the subject they are interested in, in English — "IBM and Motorola," for instance, or "recent developments in personal computers" — in a text field labeled "Look for documents about." They then drag icons representing possible sources, local or remote, into another field.

When the query is run, the Macintosh-based front end encodes the search string according to the WAIS protocol and passes it to the specified servers. Each server translates the query into its own language, locates matching articles and returns the results to the front end.

The WAIStation application then displays headlines for each article; the citations are ranked according to probable relevance, based on algorithms that consider the position, frequency and proximity of desired terms within the text.

By double-clicking on the headline, users can get the full text of any of the articles. And if the user drags the most useful titles into a bin labeled "Similar to" and reruns the search, the system will track down additional articles that share a large number of words with those selected.

Peat Marwick completed its WAIStation testing in February. In part because the cost of maintaining a real-time wide-area link among its many offices would be "substantial," according to Palmer, the company has not made a commitment to the system and is still considering a variety of alternatives. But, he said, "we are still extremely interested in the WAIS concepts. It's a most promising technology." — By Henry Norr

The WAIS system has three components:

▶ **Server software.** Any information source capable of locating and presenting text in response to a request in WAIS format can function as a server; the source can be on the user's own machine, on a LAN or at a remote site connected by modem. The WAIS client software can keep track of multiple servers, search any or all in response to a single request and consolidate the results.

Thinking Machines now includes the WAIS text-indexing and retrieval software free with its Connection Machines, a line of massively parallel systems that range in price from $100,000 to $5 million, according to Kahle. In addition, the companies participating in the project developed a sample server that runs on standard Unix systems. But any text-retrieval program on any platform, including the Mac, could be adapted to function as a WAIS server.

▶ **Protocol.** To foster the development of WAIS-compatible data sources, the four companies created an open protocol for transmitting queries and responses. It is based on an existing standard, the National Information Standards Organization's Z39.50 protocol, but is enhanced in several ways, such as by the addition of support for audio and video information.

▶ **Clients.** WAIS was designed to support a variety of interfaces running on various platforms and tailored to different niches.

The system does not rely on a specialized query language; the front end simply passes English-language search strings entered by the user to the server.

In addition to the prototype WAIStation interface and Apple's experimental personal newspaper, front ends already are available for the X Window System and GNU emacs, an extensible text editor that runs under a freely distributed Unix-like operating system developed at the Massachusetts Institute of Technology in Cambridge.

To promote the WAIS concept, Thinking Machines is making source code for the system available over the Internet or by mail. The code comes free of charge but without support. Using the software, programmers at MIT and elsewhere already have created more than 20 WAIS servers, including a poetry server, a weather server and a catalog of government programs. Thinking Machines will maintain a publicly accessible directory of servers, which will include descriptions of all known servers and special files that allow WAIS front ends to plug into them.